AD_____

Award Number: W81XWH-07-1-0510

TITLE: African Americans and Prostate Cancer: A Spatial and Multilevel Analysis of Post-treatment Care and Outcomes

PRINCIPAL INVESTIGATOR: Oliver, M. Norman, M.D.

CONTRACTING ORGANIZATION: University of Virginia
                                                    Charlottesville, VA 22904

REPORT DATE: August 2008

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                         Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
                                           Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1 Aug 2008	Annual	12 Aug 2007 – 31 Jul 2008

4. TITLE AND SUBTITLE

African Americans and Prostate Cancer: A Spatial and Multilevel Analysis of Post-treatment Care and Outcomes

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-07-1-0510

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Oliver, M. Norman, M.D.

George J Stukenborg, Ph.D., Kristen M Wells, MPH

E-Mail: noliver@virginia.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Virginia
Charlottesville, VA 22904

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report covers the first year of a two year effort to assess racial disparities in the post-treatment care and outcomes of prostate cancer. During this reporting period, the research team successfully acquired the SEER-Medicare database files, including the restricted census tract variable, for incident prostate cancer cases diagnosed between 1995 and 2002 for all 14 SEER cancer registries. From this data file, 2 geocoded prostate cancer data sets were developed: one with all cases in the SEER-Medicare data set, and a second that included only cases where the patient was 66 years of age or older at the time of diagnosis. A geographical information system (GIS) was developed using the cartographic boundary files and area-level demographic and socioeconomic measures data files available through the U.S. Census Bureau. The research team is currently integrating the spatially-referenced variables from the SEER-Medicare prostate cancer data sets into the GIS, and is preparing the data for statistical spatial clustering analyses and multilevel modeling.

15. SUBJECT TERMS
Prostate cancer, GIS, spatial analysis, SEER-Medicare

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	7	**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U			

Table of Contents

1. INTRODUCTION

Racial and ethnic disparities in prostate health and prostate cancer treatment and follow-up exist, and the reasons for such disparities are unclear. Variations in dietary fat intake, exposure to environmental hazards, physical activity, access to and use of health care services, and genetic susceptibility have been proposed as possible explanations. Additionally, it is well established that prostate cancer exhibits a striking degree of geographic variation in its patterns of incidence, morbidity, and mortality at the local, national, and international level; therefore, other measured and unmeasured characteristics of place, including neighborhood characteristics such as socioeconomic status, physical environment, availability of municiple services, and local political and cultural characteristics, need to be considered. While previous studies have commonly utilized variables at the census tract level as proxy measures of exposure for individuals, this study conceptualizes area data as being characteristics of place and not as surrogate measures for individuals. The proposed study will utilize patient location as a random variable in a hierarchical regression modelling approach to examine how race and ethnicity are related to differences in prostate cancer health, treatment types, and differences in patterns of care post-treatment. The specific aims of this study are: to utilize GIS technologies to map incident cases of prostate cancer to U.S. Census tracts; to investigate the independent effect of race and ethnicity on prostate cancer treament while adjusting for patient-level and area-level characteristics, to investigate the independent effect of race and ethnicity on post-treatment PSA surveillance while adjusting for patient-level and area-level characteristics, and to investigate the independent effect of race and ethnicity on the probability of recurrent prostate cancer among patients receiving definitive therapy while adjusting for patient-level and area-level characteristics.

2. BODY

2.1. Personnel

As presented in the initial grant application, the Principal Investigator (PI) for this study is M. Norman Oliver, MD., M.A., and the co-investigator is George J. Stukenborg, Ph.D., M.A.

In December 2007, the project PI hired Kristen Wells, MPH, as a research associate for this study. Ms. Wells' primary responsibilities have included ensuring the acquisition of the SEER-Medicare data set, the development and construction of the prostate cancer geocoded data set, and the GIS database development. Under the direction of the PI and the co-investigator, Ms. Wells is currently conducting exploratory spatial data analysis. Additionally, Ms. Wells is expanding her knowledge of the SaTScan software, which will be used to identify and evaluate the statistical significance of local clustering.

The research team meets on a regular basis to discuss study progress and to address issues of concern and challenges related to the study.

2.2. Task 1: Identify and construct geocoded data sets (Months 1-12)

2.2.1. Data Acquisition

Incident prostate cancer cases with a diagnosis between 1995 and 2002 were obtained from the Surveillance, Epidemiology, and End Results (SEER) Program, a population-based cancer registry of the National Cancer Institute (NCI). A formal request to obtain the SEER-Medicare-linked data set for the 14 SEER registries was received by the SEER Program on January 14, 2008, and formal approval for the release of the non-restricted SEER-Medicare database variables was granted by the SEER Program on January 16. A formal request for the release of the restricted SEER-Medicare database variables was sent to each of the SEER PIs on January 24, and requests for supplemental supporting documents and data were handled over the following month; approval from the PIs occurred on a rolling basis between February 29 and April 24, 2008. Approval for the release of all requested restricted variables (patient census tract, unencrypted physician identifier number, and unencrypted hospital identifier number) was granted by 13 out of 14 SEER registries; the Greater California registry approved the release of only the patient census tract variable. The data disks were received at the University of Virginia on June 15, 2008.

2.2.2. Geocoded prostate cancer data set development

Individual-level characteristics for prostate cancer cases were obtained from the Patient Entitlement and Diagnosis Summary File (PEDSF), which is the SEER data file component of the SEER-Medicare linked data set, as well as several types of Medicare files included in the linkage. Variables obtained from the PEDSF file include: cancer site and histology, tumor behavior, month and year of diagnosis, age at diagnosis, histological stage, cancer-directed surgery, grade, diagnostic confirmation, site-specific surgery, treatment by radiation, extent of disease codes for prostate pathology, date of death, and ICD code for cause of death (where applicable). Variables obtained from the Medicare files include: demographics, event dates, secondary diagnoses, procedures from Part A hospitalizations, physician claims, dates of service, diagnosis and procedure codes, facility provider numbers, revenue center codes, and beneficiary demographic information.

Upon receipt of the data disks, descriptive analyses of study-related variables were performed, data sets were cleaned, and, where applicable, merged. The merging and linking of data files was performed via each case's unique patient identifier.

The full PEDSF 1995-2002 prostate cancer data file contained 220,390 records. A valid 1970/1980/1990 census tract was available for 177,946 (80.5%) of the cases. For 99.9% of the cases where a valid 1970/1980/1990 US census tract was not available, a valid 2000 census tract was provided. A valid census tract was not available for 74 (0.1%) of the cases.

A reduced PEDSF data set was created that contained all cases age 66 and over at the time of diagnosis; a total of 163,849 cases met this criterion. Of these, a valid 1970/1980/1990 census tract was available for 107,395 (65.6%). For 100% of the cases where a valid 1970/1980/1990 US census tract was not available, a valid 2000 census tract was provided. A valid census tract was not available for 25 (.01%) of the cases.

Using the census tract assignments provided in the SEER-Medicare data set, the research staff is currently mapping the census tract location of each of the cases.

2.3. Task 2: Integrate data to build GIS and conduct exploratory spatial data analysis (Months 10-18)

2.3.1. GIS Development

Cartographic boundary files were downloaded from the U.S. Census Bureau Web site (http://www.census.gov/geo/www/cob/index.html) in ArcView Shapefile (.shp) format. All boundary files were obtained from the Census Bureau's TIGER geographic database and were designed specifically for use in GIS mapping applications.

Area-level demographic and socioeconomic measures were obtained from the U.S. Bureau of the Census Summary Tape File 3A from 1990. U.S. Census Bureau data sets were downloaded at the census tract level from the U.S. Census Bureau Web site (http://factfinder.census.gov/home/saff/main.html?_lang=en). Files included: percent of population living below the federal poverty level, percent living at 100-200% of poverty, percent with rural status, percent with less than a high school education, percent with at least 4 years of college, and median household income.

Compatibility of different spatial coverages was be addressed by using a common coordinate projection system for all spatially-referenced data; the Universal Transverse Mercator (UTM) coordinate system was utilized for this study. Latitude and longitude coordinates were stated in decimal degrees and the North American Datum 1983 (NAD83), the official datum used for the primary geodetic network in North America, was used.

2.4. Task 3: Develop multilevel regression models (months 12-24)

The relationship between race and ethnicity and differences in prostate cancer treatment type and post-treatment care will be assessed via a series of multilevel logistic regression models with random effects. The development of such models is slated to begin after the completion of the GIS.

3. KEY RESEARCH ACCOMPLISHMENTS

Key research accomplishments during the 12 months of this 24- month grant cycle include:

- The hiring of a research associate who specializes in the use of GIS for the analysis of cancer data
- Successful acquisition of SEER-Medicare 1995-2002 data set, including the release of all requested restricted variables from 13 out of 14 SEER registries and the release of the census tract variable from the fourteenth registry.
- Development of a geocoded prostate cancer data set which contains all men aged 65 and older who were diagnosed with prostate cancer between 1995 and 2002 and have a census tract location assigned to the residential address at the time of diagnosis
- Significant progress in the development of the study GIS, including the incorporation of census tract boundaries and area-level demographic variables

4. REPORTABLE OUTCOMES

Consistent with the timeline proposed in the original grant application, the analysis of the data for this project is currently in the initial exploratory stages. The research team is currently preparing the data for higher-level analysis, and plans to submit resulting abstracts and manuscripts for inclusion in peer-reviewed scientific journals.

5. CONCLUSIONS

During the first year of this two-year study, the research team successfully acquired the SEER-Medicare data set, including the restricted census tract variable, from all 14 SEER registries, and constructed two geocoded prostate cancer data sets for use in spatial and multilevel analyses. Consistent with the timeline proposed in the original grant application, the research team is currently integrating spatially referenced data into the constructed GIS, and is conducting exploratory spatial data analyses.